

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://SPIDigitalLibrary.org/conference-proceedings-of-spie)

## Real-time image and video processing for advanced services on-board vehicles for passenger transport

Sergio Saponara

Sergio Saponara, "Real-time image and video processing for advanced services on-board vehicles for passenger transport," Proc. SPIE 10670, Real-Time Image and Video Processing 2018, 106700I (14 May 2018); doi: 10.1117/12.2303918

**SPIE.**

Event: SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, Florida, United States

# Real-time Image and Video Processing for Advanced Services on-board Vehicles for Passenger Transport

Sergio Saponara

<sup>a</sup>Dept. of Information Engineering, University of Pisa, via G. Caruso 16, 56122, Pisa, I,  
[sergio.saponara@iet.unipi.it](mailto:sergio.saponara@iet.unipi.it), tel/fax: +39 050 217602/522

## ABSTRACT

The paper exploits the video camera available on-board vehicles for public transport, such as trains, coaches, ferryboats, and so on, to implement advanced services for the passengers. The idea is implementing not only surveillance systems, but also passenger services such as: people counting, smoke and/or fire alarm, automatic climate control, e-ticketing. For each wagon, an embedded acquisition and processing unit is used, which is composed by a video multiplexer, and by an image/video signal processor that implements in real-time algorithms for advanced services such as: smoke detection, to give an early alarm in case of a fire, or people detection for people counting, or fatigue detection for the driver. The alarm is then transmitted to the train information system, to be displayed for passengers or the crew staff.

**Keywords:** Public Passenger Transport, Video and Image Signal Processing, Digital Signal Processors, Smart Camera Systems, Real-Time Alarm Generation

## 1. INTRODUCTION ON SMART CAMERAS ON-BOARD PUBLIC TRANSPORT VEHICLES

On-board new public transport vehicles, such as bus, coaches and trains, usually a network of video cameras is installed to provide surveillance services [1-3]. Depending in the size of a wagon (e.g. from 10 to 14 meters for a bus, around 26 meters for a train wagon) one or more videocameras are installed, see for examples Figs. 1 and 2. In Fig. 2 parts of the wagon area are overlapped, i.e. a certain zone is covered by at least 2 cameras. Therefore, there is the possibility to exploit sensor fusion to analyze the same image from different point of views. Fig. 1 is the typical situation of low cost passenger vehicles, while Fig. 2 refers to high end models or revamping of existing wagons.

Video cameras to be installed on-board wagons of passenger transport vehicles are typically operating in the visible spectrum, and according to the specific vehicle type and model, they are connected through a high-speed packet switched digital bus to a remote storage server, or using LVDS (Low Voltage Differential Signalling). Recently, high-speed wireless connections for on-board cameras have been proposed in [4-6], although in the market typically a cabled solution is preferred for both power and data transfer.

In the market both CCD (Charge Coupled Devices) [7] imaging sensors and CMOS imaging sensors are available. CCD sensors lead to a better image quality in low illumination environments, while CMOS cameras ensure the integration of the read-out circuitry in the same chip with the sensor. For example, the device in [7] uses a 1/3-inch CCD sensor with an output digital format up to SVGA resolutions, i.e. 800x600 pixels per frame. Most of cameras used in transport systems use special

lenses to achieve large fields of view (FOV), usually from 120 degrees as in [7] to 180 degrees. The output digital video frames are often coded in MJPEG or H.26X (H.264 or H.265) mode.

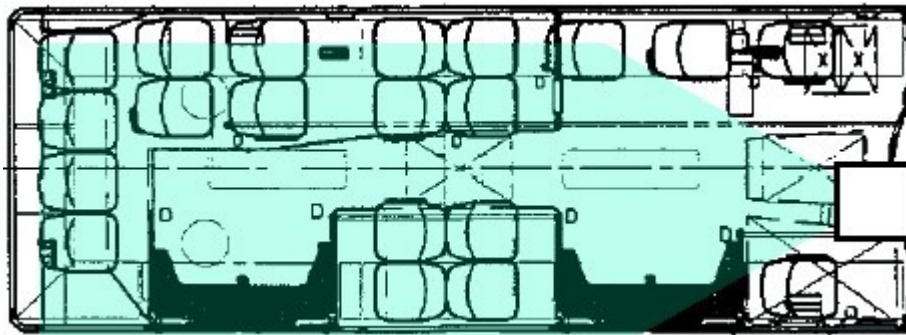


Figure 1: Example of a bus with 1 camera covering the whole wagon

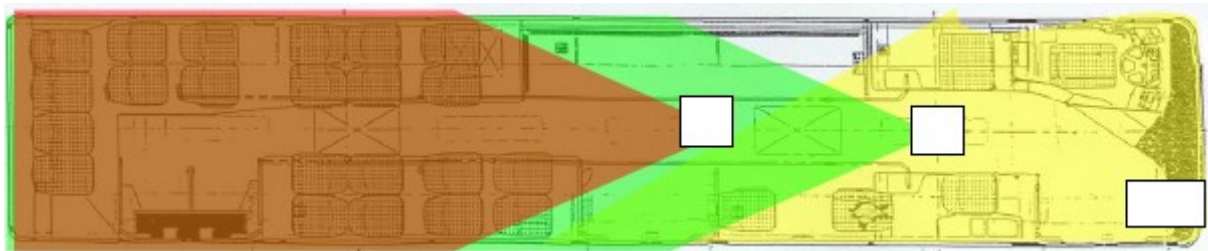


Figure 2: example of a bus with 3 cameras covering the whole wagon, some areas are covered by different view-points of multiple cameras (at least 2)

As far as the use of infrared cameras is concerned, there are two possible ways:

- use of Near InfraRed (NIR) sensors that are typically obtained at low cost by exploiting the sensitivity at NIR wavelengths of visible-spectrum devices, equipped with filtering lenses
- use of Long Wave InfraRed (LWIR) sensors operating at wavelength from  $\mu\text{m}$  7 to 14  $\mu\text{m}$ . For example, the ThermiRail in [8] is a LWIR camera that generates H.264-coded videos, with a resolution of 640x480 pixels/frame (VGA) at a frame rate up to 30 frames per second (fps).

Typically the LWIR solution is more costly than a simpler CCD or CMOS detector and is still a niche market for on-board wagon installation, for public transport systems.

However, at the state-of-the-art, the video network on-board public transport vehicles (i.e. trains, coaches, bus, ferryboats) is typically used in a “passive way”, recording the scenes inside the wagon but without any advanced service of recognition or detection [5, 9, 10]. Instead, this paper proposes an intelligent uses of such cameras to provide advanced services on board trains.

The rest of the paper is organized as follows. Section II reviews smart systems that have been proposed at the state-of-art to implement advanced passenger services. Section III presents the system level architecture. Section IV deals with an early detection smoke and fire system, from both an algorithmic and hardware point of view. Section V discusses people counting services that can be integrated in the same embedded platform. Section VII makes a comparison to state of art and draws some conclusions.

## 2. SMART CAMERA BASED SYSTEMS FOR ON-BOARD SERVICES

The aim of this work is exploiting the cameras already installed on-board wagons to implement new services for the passengers. Among the many possible services that can be implemented there is the increase of the passenger safety, by using the cameras for smoke detection and early warning in case of fire. This function is important to reduce the alarm generation latency. Another possibility is implementing automatic people counting systems, by adopting proper recognition algorithms. This function can be useful to enable services like e-ticketing or automatic climate control. Furthermore, the camera installed in front of the cab can be used to implement a detection system for driver's fatigue and drowsiness.

In the state of art, the use of LWIR cameras plus a detection processing board has been tested for fire alarm generation by implementing process algorithms that detect the thermal change vs. background temperature when there is a fire. Another possible function is people detection (and then counting) by detecting the thermal change vs. background temperature, around 18 - 20 °C in a climatized wagon, due to a person which has a body temperature at about 36-37 °C. However, such systems at the state of art have been proposed using the ThermiRail camera, discussed in Section 1, which has an horizontal FoV of 90°. To provide an all-around view, at least 4 cameras, each limited to 90° FoV, have to be used thus increasing by 4 times the cost of each installation. Moreover, the cost of a LWIR thermo-camera is one order of magnitude higher than the cost of a video camera operating in the visible spectrum.

When detecting people, thermo-cameras work very well during night, with background temperature below 36 C°, whereas they may fail in case the background temperature is near 36 C° (e.g. an hot day in summer). For fire detection, thermo-cameras work well when the fire is started and the temperature is high. Hence, fire alarm systems relying on the use of thermo-cameras may suffer of high latency. The fact that thermo-cameras work better than visible ones in dark environments is not a big advantage, since passenger wagons are usually illuminated. Moreover, CCD cameras work well also in low light conditions. Indeed, the CCTV camera in [6] has a sensitivity of 0.02 Lux.

With respect to the use of LWIR cameras, this work aims at reusing the already installed camera system, as shown in Fig. 1 (single camera system) or Fig. 2 (multiple camera system). This way there is no overhead in terms of installations costs and purchasing costs of expensive LWIR imaging sensors.

For the smoke detection and fire alarm generation, this work proposes to detect, thanks to real-time image processing, the presence of smoke, rather than measuring the temperature increase due to the fire. This way, an early alarm generation can be obtained with a detection latency reduced to few seconds. Instead, the fire detection latency is several minutes in state of art systems detecting temperature changes and/or volatile chemical substances in case of a fire through temperature and chemical sensors.

Moreover, by a proposed fusion of data coming from the already installed visible cameras, and the data coming from other sensors already installed on-board trains, e.g. temperature resistive sensors or fire chemical sensors, or door bars [11] for people counting, the false alarm rate and the missed detection rate can be reduced.

Therefore, thanks to data fusion, the system redundancy and the system accuracy can be increased, but with minimal cost and installation overheads. Indeed, this work does not propose to install new devices, but just exploiting already installed on-board CCD or CMOS videocameras.

This is the case of old trains and buses. Instead, in new ones where often multiple cameras are installed, see Fig. 2, we can exploit the fact that the same area is covered by different video cameras,

but with different views, to reduce false alarms rate or missed detection rate. Indeed, an alarm is raised only if the same detection is done by both video cameras covering the same zone. Moreover, the same infrastructure (video cameras, networking devices, acquisition and video processing units) can be used to implement multiple services (e.g. smoke/fire alarm, people counting, smoke detection in smoke-forbidden zones).

### **3. SYSTEM-LEVEL ARCHITECTURE OF THE ON-BOARD VIDEO NETWORK FOR ADVANCED SERVICES TO THE PASSENGERS**

The proposed work aims at reusing the on-board video camera network, using commercial cameras operating in the visible spectrum, to improve some functional and safety aspects of a train. The proposed approach does not require installations of extra sensors and hence it minimizes purchasing and installation costs. A simplified and common on-board video camera network is composed by many IP cameras per wagon, usually from 2 to 8. Usually, a local Switch connect the video streams to a router that redirects the data to the main communication network, to the digital video recording (DVR) sub-system, to the service terminals and to other functional servers for fire protection and HVAC (Heating, Ventilating and Air Conditioning). Indeed, the videocamera system is a source of streaming videos, which are transferred towards a passive storage information system. Adding an embedded node for acquisition and processing of the video streams, anomalies can be detected and new information can be sent to each system connected on the network.

The main idea is that in all video acquisition and processing nodes the acquisition is done on the digital domain. The acquisition and processing node can be implemented as a low-cost embedded system with low-power budget (limited to few Watts) thus easing its installation in a service rack. The design of the video acquisition and processing node in this work has been done considering the multiplexing of at maximum 4 camera streams to a single node. As consequence, for each wagon, having from 1 to 8 video cameras, the number of video acquisition and processing node installed is from 1 to 2.

Concerning the frame rate to be used for video acquisition, according to the guidelines of the American Public Transportation Association (APTA) [12], the minimum is 4 fps for PAL-based systems for low-traffic areas, or areas where only walking-pace motion is likely. In case videocameras are covering trackside operations or areas where fast-moving objects are likely to be observed, then at least 12 fps should be used. In case videocameras are covering passenger areas, or areas containing any form of emergency button, a two-speed capability must be incorporated into the cameras' codec to provide 4 fps in normal mode, and 12 fps in emergency mode.

A standard camera such as that in [7], with SVGA resolution (800x600 pixels/frame) and a frame rate of 12 fps, produces an uncoded data stream of about 5.76 Mpixel/s. Since the output is H.264 coded, then the output bit-rate is below 1 Mbps. In this scenario, a set of 4 cameras connected on the same video switch requires about 4 Mbps for the transmission. The use of a standard PoE (Power over Ethernet) switch at 100 Mbps ensures the technology feasibility of the network, even if more information has to be used or a high video quality is required. The local switch is connected to the local image acquisition and processing node, which acquires the video stream and processes it as an embedded node, delivering only the necessary information on the network.

The additional data rate to be sustained by the acquisition and processing node depends on the information transmitted. To transmit the smoke detection preview (if the video stream has to be visualized on a terminal) the cost to be paid is the same as a video acquisition. The train conductor, in this case, can activate the alarm, increasing the reaction time for fire prevention.

Otherwise, in case the video-based system is automatically generating the alarm, the information transmitted to other prevention and sensing systems is negligible in bandwidth. The best solution is providing both an automatic alarm, to reduce the latency of the warning system, and at the same time transmitting the sequences that triggered the alarm so, in case of a false alarm, the train staff can stop the alarm.

Instead, in case of people-counting techniques the cost of the data payload amounts to few kbps, since only numbers (i.e. how many people) and wagon information have to be transmitted. The people-counting system provides to the train staff the information about how many people are on-board each wagon. This information can be used for multiple purposes. To name just a few we can consider (i) giving a feedback that can be used by the climate control system since the number of people influences the humidity and temperature levels in the wagon; (ii) comparing this number with the number of sold tickets as a countermeasure vs. non-paying passengers; (iii) avoiding that some wagons are crowded, while other wagons are almost empty, particularly for trains where the seat is not assigned when buying the ticket (e.g. local trains, tram).

The feasibility of the proposed improvements guarantees the reuse of the existing on-board video camera network and of the networking infrastructure. The hardware modification of the system we propose only implies the insertion of the embedded acquisition and processing node. Since each of this nodes can manage a few number of real-time video streams, then it should be hosted locally to each wagon, near the local video switch. This solution provides also a safety redundancy of each node. This means that if a node crashes, or has a failure, a wagon-near node may replace its function (safety mechanism).

It has to be noted, that this manuscript refers to cabled interconnections, particularly those based on Ethernet, although the system can work also with wireless connections, e.g. Wireless Personal Area Network (WPAN) and Wireless Local Area Network (WLAN) providing bandwidth of several Mbps at distances within 10 meters and hundreds of meters, respectively. The use of WPAN and WLAN on-board train is increasing. The new IEEE802.15.4e protocol is particularly suited for safety-critical applications with harsh requirements in terms of real-time and low-latency, such as industrial and transportation fields [13].

#### 4. SMOKE AND FIRE REAL-TIME PREVENTION ALGORITHM

As a case study, this Section presents the algorithm and the implementation on a programmable system, of one of the possible features running on an embedded acquisition and processing node. Standard fire or smoke sensors are generally dislocated in few points of a wagon. The response time from a triggering of a fire event may be in the orders of minutes with those sensors. A video camera analysing system, enhanced by the acquisition and processing node that implements proper video processing algorithms, by monitoring the whole space of a wagon, can analyze the presence of smoke and recognize it faster. The probability of the smoke presence can be sent to the fire protection server as an advice of lowering and refining the thresholds of temperature and chemical sensors. Fig. 3 shows, as an example, the smoke presence probability as a Pre-Alarm signal. The red line of Fig. 3 is a dynamic pre-alarm threshold where the presence of smoke is detected with about the 75 % of accuracy during the test validation, see Eq. (1).

In Eq. (1) TP are the true-positive results of the elaboration, TN the true-negative, FP the false-positive and the FN the false-negative.  $Accuracy = (TP+TN) / (TP+TN+FP+FN)$  (1)

Unfortunately, an accuracy of 75 % is low for fully autonomous system. From the analysis of the results of the proposed technique, applied to real on-board train scenes, it emerges that the limit of accuracy is mainly due to the FP parameter. This is why in previous Section, we suggested in case of an alarm transmitting also the sequences that generated the alarm to the train crew staff. The staff people can stop the alarm in case it is false. Therefore, the proposed system is an assistance system, rather than a completely autonomous one.

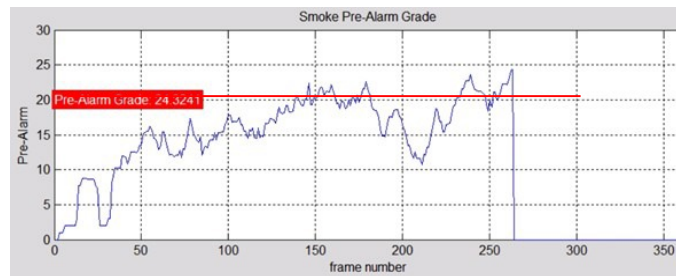


Figure 3: Smoke presence probability as a Pre-Alarm signal

From the triggering of a fire event to the first pre-alarm occurrence, the video smoke detection system takes from 10 to 20 seconds of latency in the tested video set. A fine-tuning of the algorithm parameters and an extensive dataset for learning and validate cases, in on-going to improve the accuracy of the detection. Alternatively, if the train conductor wants to monitor the smoke presence, the elaborated video is visualized on a connected terminal.

For example, Fig. 4 shows a scene in a container where a chair is on fire and the smoke is rising up. The scene shows a worst case since the fire is not directly seen by the video camera, due to an occlusion, but the alarm is still generated by detecting (green blobs) the smoke that is rising up. The scene in Fig. 3 has a gray background. However, the proposed technique is able to distinguish the gray smoke from the gray background. The highlighted green blobs in Fig. 4 are the more credible candidate areas of smoke presence. De facto green blobs indicate the ROI (region of interest) where the algorithm focuses the processing to discover smoke.



Figure 4: Smoke presence visualization on a terminal

The algorithm used for the smoke recognition follows the video processing flow showed in Fig. 5, similarly to what we discussed in [2]. The first step in the algorithm is the video decoding. The



acquired H.264 stream is converted to an RGB frame sequence for a more suitable elaboration. Motion segmentation, using background subtraction, provides a logical information of moving pixels. Color segmentation extracts pixel belonging to a specific gray shade range, the possible smoke's color. Next step is the merge of the two logical image segments, a simple pixel to pixel logical AND is used. A morphological filter removes and reduces many type of 2D noise in the merged image. The logical image is then labeled, assigning a unique *blob id* for each connected set of positive pixels. Each blob, or a more significant subset, is then analyzed frame by frame to extract static and dynamic features. The next steps of the algorithm in Fig. 4 for the features analysis are designed to be modular and parametrical. It means that every sub-processing block can be substituted or at least removed to fit in different hardware or software solution. In addition, a high number of parameters and thresholds can be refined or tuned to arrange the flow into different scenarios. Static features are coming by a single frame processing, they include:

- Blob Area, the number of pixel composing the blob.
- Blob Turbulence, a measure of the ratio between the area and the perimeter of the blob.
- Bounding Box Coverage, the percentage of the blob's pixel inside the blob's rectangular bounding box.
- Bounding Box Aspect Ratio, the ratio between the width and the height of the bounding box.
- Gray Standard Deviation, the grayscale standard deviation value of the blob's pixel.

The most credible blobs, representing a possible smoke event, are taken into account. Small area, regular or thin shape blobs are discarded in the algorithm flow to speed up the next dynamic analysis. The selected blobs are the inputs of the dynamic features extraction. Dynamic features are coming by the analysis of a sequence of frames; in this case, a window of 2 seconds allows for good results. Dynamic features include:

- Blob Tracking; estimates the blob motion direction. It is implemented using the sum of absolute differences (SAD) technique in the surround of the original blob position. Smoke's blobs move toward the top of the scenario, not down.
- Edge Occurrence, is the analysis of the possible incoming gradient discontinuity inside the tracked blob.
- Dynamic Blobs Displacements, is the 2D distribution analysis of the blobs, during the 2 seconds window. If the time-space distribution of candidate blobs increase in a certain point it increase the likelihood of having a real smoke event.

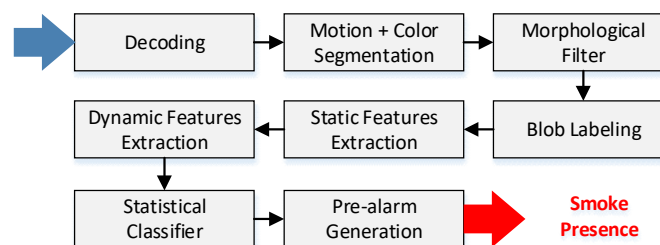


Figure 5: Smoke detection algorithm

The whole set of features provides a descriptor of each tracked blob, the statistical classifier selects those who represent the possible “smoky” blobs. The higher is the number of recognized blobs and the likelihood of the dynamic features recognition, the higher is the pre-alarm output.



## 5. FIRE PREVENTION IMPLEMENTATION

Suppose to process a H.264 video stream, at the output of the VPort 16-M12 camera in [7] (800x600 pixels) using 12 fps as suggested by APTA in [12]. The data rate of the encoded stream can be estimated using Eq. (2), suggested as heuristic formula in [14] to keep high visual quality. In Eq. (2) the image width and height is expressed in pixels, and the motion rank is an integer between 1 and 4, 1 being low motion (e.g. a static scene), and 4 being high motion (e.g. a sport scene with lots of zooming, change of scene etc.). In [14] the motion rank is the amount of image data that is changing between frames. For surveillance applications on-board trains the motion rank in Eq. (2) is typically between 1 and 2.

$$[width] \cdot [height] \cdot [fps] \cdot [motion\ rank] \cdot 0.07 = [bitrate] \quad (2)$$

Under the above assumptions from Eq. (2) the H.264 coded bit-rate will be from 400 kbps to 800 kbps, depending on the motion characteristics of the scene. A set of hardware requirements to implement the proposed system are:

- Minimum 16.5 Mbyte of RAM memory to store one second of video, i.e. an uncoded set of 12 frames, using 24 bit/pixel (color images).
- A communication data rate between 400 kbps and 800 kbps for the H.264 coded stream, which can be easily managed by a 10/100 Ethernet link or similar communication technologies available on-board trains such as PoE or Broadband Power Line.
- A processor with H.264 decoder and DSP (Digital Signal Processing) capability.

With reference to the algorithm described in Section 4 and Fig. 5, the following Fig. 6 shows the contribution of the different algorithmic parts to the computational burden. The main limiting tasks are the feature-extraction and the motion estimation. Motion estimation doesn't include only background subtraction but it provides a set of filters and special computation to take into account and manage some of the most common visual side effect in a train scenario, e.g. fast light change when train go through tunnels, fixed region of minor interest as windows and so on. Static and dynamic features extraction can be reduced in complexity setting the *max\_number\_of\_blob* parameter. The data and the implementation presented in this paper refer to the default algorithm where the parameter *max\_number\_of\_blob* is equal to 100. Heuristic analysis shows an unacceptable performance degradation with the parameter value less than 40.

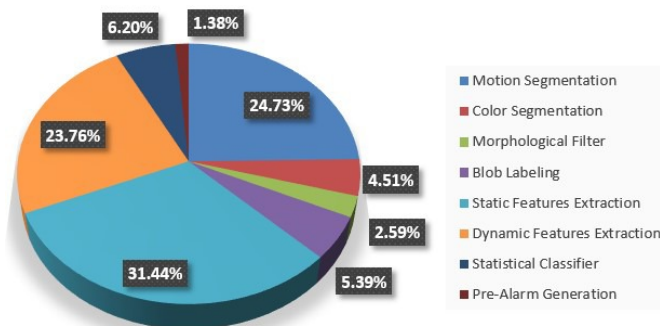


Figure 6: Algorithm complexity per task

A custom board including a video DSP or a video SoC (System-on-Chip) with sufficient external RAM can be designed to implement the embedded acquisition and processing node. In order to minimize the hardware design effort, a COTS (Commercial Off the Shelf) solution is taken into

account. For example, a Raspberry PI board has been used for the evaluation tests. To reduce costs, in case of a large volume production, a custom system can be implemented. For example, a printed circuit board (PCB) of about 10 cm per side can host a TMS320DM368 low-power DSP from TI [15]. This is a high-performance Digital Media processor equipped with a 32-bit ARM926EJ core, with DSP instruction extension and single-clock latency MAC (multiply and accumulate), running with a clock frequency up to 432 MHz, with single-chip integrated hardware co-processors to implement image processing tasks such H.264/MJPEG encoded/decoding, lens distortion correction, or face detection. The TMS320DM368 has on-board 16 kbyte instruction cache, 8 kbyte data cache, 32 kbyte RAM and 16 kbyte ROM, plus a rich set of integrated controllers towards external memories (Nand Flash, DDR2 dynamic RAM, Secure digital memory card) and I/O interfaces for high-speed data links (10/100 Ethernet, USB) and for control links (SPI, UART, I2C).

If multiple cameras should be managed by a single, e.g. 4 as discussed above, the abovementioned memory storage and data-rate requirements should be multiplied by a factor of 4 becoming 66 Mbyte of RAM memory to store one second of video.

For the processing unit, a COTS device with high computational capability should be selected. To keep low the power consumption a suitable solution is using FPSoC (Field Programmable System-on-Chip) devices, which integrate in the same chip software-programmable cores and configurable FPGA (Field Programmable Gate Array) co-processors. Indeed, literature [16] has demonstrated that for computing intensive tasks FPGAs offer a better trade-off between computational capability and power consumption than software-programmable DSP or General Purpose Processors (GPP) or Graphic Processing Units (GPU). The design of ASIC (Application Specific Integrated Circuit) will entail too high design time and cost and is not justified by a market, which is not yet a large volume one. For example, a COTS device that can be selected for the target applications of this paper is the ZynQ-7000 FPSoC. It features two ARM Cortex-A9 processors, with a maximum clock frequency of 1 GHz and more than 800 kbytes of memory (L1 cache for each processor, shared L2 cache and memory buffer). Each programmable core is enhanced by NEON & Single/Double Precision Floating Point mathematical coprocessors, by a rich set of memory (DDR2, DDR3, Nand Flash, SD card) and I/O interface (Ethernet USB, CAN, I2C, SPI, UART, USB) controllers, and by embedded FPGA logic for coprocessor design. For example, in terms of configurable logic the ZynQ-Artix Z7030 device features 125K Xilinx logic cells corresponding to about 2 Million equivalent logic gates, 9.3 Mb of RAM blocks, 400 DSP slices each equipped with a 18x25 Multiply and accumulate hardware unit. The computational capability of Z7020 is about 600 GMACs. With reference to the algorithm in Fig. 5, and using the VPort 16-M12 [7] device as input source, an acquisition and video processing node, based on the ZynQ platform device, can manage at least 2 video-cameras. To manage 4 cameras we should use more powerful processing unit, such as the Tegra GPU from NVIDIA.

## 6. PEOPLE-COUNTING TECHNIQUES

The algorithm for people counting requires less complexity than the algorithms for image-based smoke and/or fire alarm detection. Some partial outputs of the processing chain in Fig. 5 are reused. Motion segmentation, one of the high complexity steps in the chain, is a sub-process shared between smoke detection and people counting algorithms. After that, blob labeling and blob tracking strategies are used. The tracked blobs are geometrically analyzed and a human shape classifier provides the estimated people count. To this aim, we can exploit co-processors already integrated in

the selected platforms, such as the face detection and hand tracking features [17] of the TMS320DM368.

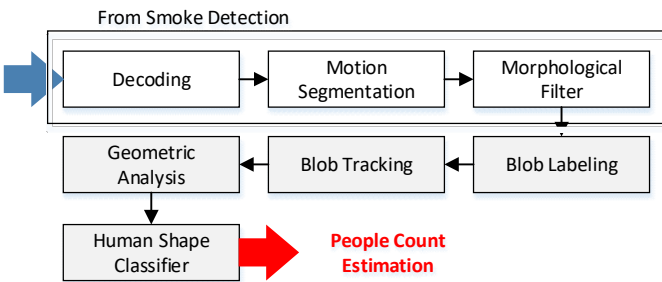


Figure 7: People counting algorithm

Fig. 7 shows the algorithm chain for people counting. Its implementation is done in real-time on the Raspberry PI board for a single camera input and on the ZynQ FPSoC for multiple-camera inputs. The people counting technique can be used for e-ticketing crosschecking, transmitting to the central station the total number of passenger in the train. In addition, the climate control system in each wagon can be tuned and regulated as function of wagon passenger number. This allows a power consumption optimization for the climate-control and a comfortably dynamic climate adaptation. It is worth noting that the camera based counting people system can provide a better redundancy vs. conventional counting systems using intelligent doors, as in [11]. From the proposed analysis the accuracy of camera based counting system is higher than 90 %, aligned with results published at the state-of-the-art [18].

By fusing the data coming from the two systems (intelligent doors, as in [11], and video cameras, as proposed in this paper) the accuracy of the detection system, see definition in Eq. (1), can be further improved.

## 7. STATE OF ART ANALYSIS AND CONCLUSION

### 7.1 State of art comparison

To be noted that using data fusion, and re-using the same videocamera infrastructure for multiple additional services for passenger public transport, such as fire alarm, smoke alarm, people counting, and so on, is missing in other works at the state-of-the-art. For example the work in [19], considered just fire alarm generation in rolling stocks, while using video sensors multiple sensors can be implemented. Instead, using as in [2] only 1 camera the achieved rate of missed detections or false alarms in not acceptable for a fully autonomous system, since the accuracy is 75%, and the intervention of the staff crew is required. By using multiple cameras covering the same zone, as proposed in this paper, the false alarm rate and missed detection are improved, and can go above 90% This puts the basis for an autonomous system. Table 1 summarises the comparison to the stat of art.

Table 1: State of art comparison

Work	Multiple services	Fusion of multiple views
[19]	No, only fire alarm	No
[2]	Yes, fire alarms, people counting, driver	No
This work	drowsiness detection	Yes

## 7.2 Conclusion

The paper has presented a video-camera networking system to implement advanced services on board public transport vehicles, such as trains, coaches, ferryboats. State of art transport systems typically adopt video-cameras just for surveillance systems, analyzed off-line. Instead, in the proposed work the idea is that for each wagon, an embedded acquisition and processing unit is used, which is composed by a video multiplexer, and by an image/video signal processor that implements in real-time algorithms for advanced services such as: smoke detection, to give an early alarm in case of a fire, or people detection for people counting, or fatigue detection for the driver, automatic climate control. The alarm is then transmitted to the train information system, to be displayed for passengers or the crew staff.

## REFERENCES

- [1] Mahasak Ketcham “The intruder detection system for rapid transit using CCTV surveillance based on histogram shapes”, IEEE Int. Joint Conf. on Computer Science and Software Engineering (JCSSE), pp. 1-6, 2014
- [2] S. Saponara, L. Pilato, L. Fanucci, “Exploiting CCTV camera system for advanced passenger services on-board trains”, 2016 IEEE International Smart Cities Conference (ISC2), pp. 1 – 6
- [3] B. Aldred, “London Underground Platform to Train CCTV Studies”, IEE WC2R, Savoy Place, London, UK, 1997, pp. 1-12
- [4] S. Saponara, F. Giannetti, B. Neri, G. Anastasi, “Exploiting mm-Wave Communications to Boost the Performance of Industrial Wireless Networks”, IEEE Transactions on Industrial Informatics, vol. 13, n. 3, 2017, pp. 1460 – 1470
- [5] Hyunjeong Jo et al., “Wireless security method for on-board centered train control system”, IEEE ICCCT, pp. 88-93, Dec. 2012
- [6] S. Saponara, F. Giannetti, B. Neri, “Design exploration for millimeter-wave short-range industrial wireless communications”, IEEE IECON 2016, pp. 6038 – 6043
- [7] Moxa Inc., “Onboard CCTV Surveillance with Rear-Facing IP Cameras”, [http://www.moxa.com/Solutions/Railway/pdf/CCTV\\_success1.pdf](http://www.moxa.com/Solutions/Railway/pdf/CCTV_success1.pdf), accessed 4<sup>th</sup> July 2016
- [8] FLIR Inc, “Intelligent transport systems, detection and monitoring solutions for traffic and public transportation applications”, pp. 1-11. N. IT\_004\_EN, 2016
- [9] I. Salaberria et al., “Towards a train-to-ground and intra-wagon communications solution capable of providing on trip customized digital services for passengers”, Lect. Notes in Computer Sc., vol.8276, pp334-341,2013
- [10] G. Lisanti et al., “A Multi-Camera Image Processing and Visualization System for Train Safety Assessment”, Computer Science, Computer Vision and Pattern Recognition, July 2015
- [11] Eurotech, “DynaPCN 10-20 Automatic passenger Counter”, doc. ETH\_DynaPCN 10-20\_DS-04/2013
- [12] American Public Transportation Association, “Selection of Cameras, Digital Recording Systems, Digital High-Speed Networks and Trainlines for Use in Transit-Related CCTV Systems”, IT-CCTV-RP-001-11 Published June 2011 CCTV Standards Working Group
- [13] T. Watteyne et al., “Industrial IEEE802.15.4e networks: Performance and trade-offs”, IEEE Int. Conf. on Comm. (ICC), pp. 604-609, 2015
- [14] K. Amerasinghe, “H.264 for the rest of us”, pp. 1-26, availble from [http://www.adobe.com/content/dam/Adobe/en/devnet/video/articles/h264\\_primer/h264\\_primer.pdf](http://www.adobe.com/content/dam/Adobe/en/devnet/video/articles/h264_primer/h264_primer.pdf), accessed 4th July 2016
- [15] TI inc., “TMS320DM368 Digital Media System-on-Chip (DMSoC)”, SPRS668C, June 2011
- [16] B. Harikrishnan et al., “All programmable SOC based standalone SDR platform for researchers and academia”, IEEE First Int. Conf. on Computational Systems and Comm. (ICCSC), pp. 384-386, 2014
- [17] Dong-Ik Ko, G. Agarwal, “Gesture recognition: Enabling natural interactions with electronics”, TI white paper, SPRY199, 2012
- [18] Bin Li et al., “People counting method based on head detection and tracking”, IEEE Int. Conf. on Smart Computing, pp. 136-141, 2014
- [19] S. Saponara, L. Pilato, L. Fanucci, “Early video smoke detection system to improve fire protection in rolling stocks”, Proc. of SPIE, vol. 9139, 2014